Robust and Discriminative Image Authentication Based on Standard Model Feature

Luntian Mou[1,2], Xilin Chen[1]
[1]Key Lab. of Intell. Info. Process., ICT, CAS
[2]Graduate University of CAS
Beijing, China
{ltmou, xlchen}@jdl.ac.cn

Yonghong Tian, Tiejun Huang
National Engineering Laboratory for Video Technology
School of EE & CS, Peking University
Beijing, China
{yhtian, tjhuang}@pku.edu.cn

***Abstract*—The goal of image authentication is to accept content-preserving operations and reject content-altering manipulations. So，it is increasingly approached by extracting content-based invariant features from original images and verifying their preservation in received images at later times. Since sparsity usually implies invariance, sparse feature representation has drawn significant attention from the research community. But only if discrimination is also found with a sparse feature, can it be successfully applied in image authentication. This paper proposes a sparse feature for image authentication by exploring the biologically-motivated standard model. Experimental results demonstrate both robustness and discrimination of the feature, and its effectiveness in tamper detection and location as well.**

I. INTRODUCTION

With increasing availability of powerful image processing and manipulating tools, the credibility of image content is on the decline while the need for image authentication is on the rise. Image authentication is usually achieved through sender authentication and content integrity verification. On one hand, the actual identity of the sender and sometimes its non-repudiability should be ensured. On the other hand, the preservation of the content of an image, instead of its specific binary representation, must be confirmed. Therefore, a general principle for image authentication is to accept content-preserving operations while rejecting content-altering manipulations. Image authentication has been actively studied recently in content-based approach [1], that is, to extract content-based invariant features from original images and verifying their preservation in received images at later times. This content-based approach, known as robust hashing or perceptual hashing, is intrinsically different from the traditional data-based approach (i.e., cryptographic hashing): the former can tolerate moderate image processing while the latter is sensitive to bit-change. Such a unified content-based authentication framework (see Fig. 1) is derived from the conventional data authentication in our previous work [2]. Obviously, it is the performance of the invariant feature in use that largely determines the performance of an image authentication system.

Existing invariant features can be roughly classified into three categories: statistical features such as intensity histogram [3] and moments [4], transform domain features such as DCT [5] and Fourier–Mellin transform [6], and low-level visual features including edges [7] and feature points [8]. Although robustness is shown to some extent by these features, their discrimination abilities are not adequately addressed. Since sparsity usually implies invariance, sparse feature representation has drawn significant attention from the research community. But only if discrimination is also found with a sparse feature, can it be successfully applied in image authentication. Therefore, one such feature of sparse coding was explored by us for image authentication [2].

In this paper, we explore another sparse feature based on the standard model [9]. Contrast to sparse coding [10], which emulates only the properties of receptive fields of simple cells in the primary visual cortex (V1), the standard model is composed of hierarchical computational units simulating properties of cells along the ventral stream of visual cortex. By analyzing the characteristics of all the hierarchical features, we select one feature as a sparse representation for image authentication. Through experiments, we find that this feature exhibits excellent robustness and discrimination. Therefore, we propose the standard model feature (SMF) as a new invariant feature for image authentication.

The rest of the paper is organized as follows. Section II presents the proposed method of SMF based image authentication. The performance of the SMF is evaluated in Section III. In Section IV, conclusions are drawn and the future work is outlined.

II. IMAGE AUTHENTICATION BASED ON STANDARD MODEL FEATURE

As shown in Fig. 1, the core of image authentication is constructing a proper representation for image content based on an invariant feature. To address this core, this section brings up a feature representation and corresponding similarity measurement based on the standard model.

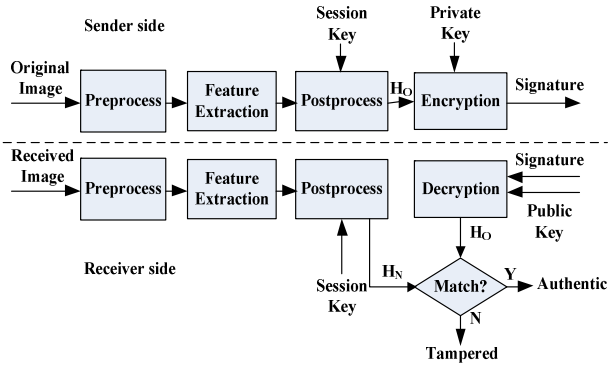

Fig. 1. A unified framework for content-based image authentication.

A. Standard Model

The standard model summarizes a core of well-accepted facts about the ventral stream in the visual cortex: 1) visual processing is hierarchical; 2) along the hierarchy, the receptive fields of the neurons and the complexity of corresponding optimal stimuli increases; 3) the initial processing of information is feedforward [11]. Feature complexity and position/scale invariance are built up along the hierarchy by alternating template matching and max pooling operations. The model is composed of at least four layers of such computational units, named *simple S* units and *complex C* units respectively. The S units combine their inputs with a bell-shaped function to increase selectivity, while the C units pool their inputs through a maximum (MAX) operation to improve invariance. The model is in accordance with several properties of cells along the ventral stream of visual cortex [11]: the first two layers correspond to V1, which contains simple (S1) and complex (C1) cells, the S2 layer corresponds to cortical area V4 or IT, and the final C2 layer is a vector of "bag of features" with global invariance by removing all position and scale information (see Fig. 2).

B. Feature Representation

The four layers of the standard model are computed hierarchically in a similar way to [12].

S1 Layer: The S1 layer is computed from the image layer via applying 2D Gabor filters. Consequently, the S1 layer has the same pyramid shape as the image layer but with multiple oriented units at each position and scale. Each unit represents the activation of a particular Gabor filter centered at that position/scale. And the response of a patch of pixels X to a particular S1 filter G is given by:

$$R(X,G)=\left|\sum X_i G_i \Big/ \sqrt{\sum X_i^2}\right|. \tag{1}$$

C1 Layer: For each orientation, the S1 pyramid is convolved with a 3D max filter, 10x10 units across in position and 2 units deep in scale. So, a C1 unit's value is the value of the maximum S1 unit that falls within the max filter. And its computation is presented as follows:

$$C1_i = \max_{\substack{postions=10\times10\\ scales=2}} (S1). \tag{2}$$

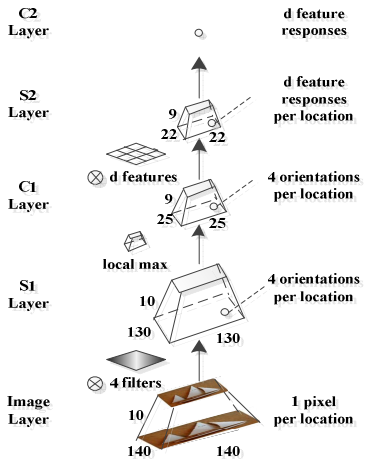

Fig. 2. Feature computation in the standard model.

S2 Layer: Template matches are performed between the patch of C1 units centered at each position/scale and each of d prototype patches, which are randomly sampled from the C1 layer of the training images. As a result, an S2 pyramid is generated with roughly the same number of positions/scales, each representing the response of the corresponding C1 patch to a specific prototype patch. The response of a C1 patch to a S2 prototype P is given by a Gaussian radial basis function:

$$R(X,P)=\exp\left(-\|X-P\|^2 \big/ 2\sigma^2\alpha\right), \tag{3}$$

where the standard deviation σ is set to 1, and α is a normalizing factor for different patch sizes.

C2 Layer: A d-dimensional vector is obtained with each component being the maximum response (anywhere in the image) to one of the d prototype patches.

Basically, the Gabor filter based S1 feature is similar with the sparse coding feature [2]. But it is high dimensional, for example, 76800D for just one scale with the size of 80×80. For the S2 feature, since at each position there are d (e.g., 4075) feature responses, it is too dense to be used in image authentication. With respect to C2, it is targeted for object class recognition and may not be informative to represent the scene of an image. Contrast to the global invariance of C2, the feature C1 bears local invariance. And it also inherits selectivity from the input layer S1. Accordingly, an assumption is made that robustness and discrimination should be owned by C1. Therefore, C1 is chosen as the standard model feature (SMF) for image authentication.

C. Similarity Measurement

Given two images, the Euclidean distance is applied in similarity evaluation. Let v and v' be the two SMF-based feature vectors representing the two images I and I', then their distance can be computed as:

$$Dist(I,I')=\sqrt{\sum_{i=1}^{n}(I_i-I_i')^2}, \tag{4}$$

where n is the dimension of the two vectors. Accordingly, the similarity between the two images is calculated as follows:

$$Sim(I, I') = \exp\left(-Dist(I, I')/med\right), \quad (5)$$

where *med* is a normalizing factor taking the value of the median of all the evaluated pairwise distances. In this way, the similarity values are normalized and uniformly distributed to the range of $[0,1]$. Thus, the similarity values between an image and its slightly modified versions are expected to be close to 1, while those of totally different images should be near 0. For a practical image authentication system, a threshold τ should be predefined so that a received image with a similarity value larger than or equal to τ is determined as authentic, or asserted as tampered otherwise.

III. Performance Evaluation

Experiments are carried out mainly over two datasets. One is a subset of the benchmarking dataset used by MPEG for evaluating image signatures. From this dataset called NOVA [13], 1000 high definition JPG images are selected, covering 10 categories from art to wildlife. Additionally, 10 super high resolution (e.g., 3072x2048) raw images [14] are also included for testing the feature's invariance to JPEG compression. The other contains 3600 video frames extracted from the training datasets of TRECVID [15]. These images from NOVA and TRECVID are used as original images, which are transformed by 6 content-preserving operations and 8 complicated manipulations respectively into modified versions to simulate received images to be authenticated. Consequently, the size of our NOVA dataset is 6020, while the other of TRECVID being 32400. The performance of SMF is compared with the sparse coding feature (SPC) proposed in [2].

A. Evaluation over NOVA

The SMF representation is computed from both original images and their slightly modified versions. And the proposed similarity measurement is applied in assessing the pairwise similarity between each original image and its modified versions for robustness evaluation, while between any two original images for discrimination evaluation. To achieve compact feature representation, the parameter settings at feature computation are tuned mainly by reducing the scales and orientations used at the C1 layer as well as the S1 layer. Accordingly, the performances of SMF are evaluated under different parameter settings, and the results are summarized in Table Ⅰ with comparison of SPC. It can be clearly figured out that SMF bears not only sufficient robustness to those content-preserving operations, but also great discrimination among different images. Since no significant performance improvement has been observed under different parameter settings, the feature C1 under 4 orientations and 1 scale is preferred as the SMF feature because it is the most compact in size (81348D under 12Ori.+11Sca., and 4075D under 4Ori.+1Sca.), and the least complex at computation. Compared with SPC (4096D), SMF also bears excellent performance of robustness and discrimination. Note that the big difference between average similarity values for different images based on SMF and SPC respectively is mainly due to the different similarity measurements employed. The performances of SMF and SPC are further demonstrated in Fig. 3 and Fig.4. It can be observed that both features achieve

TABLE I. AVERAGE SIMILARITIES EVALUATED OVER NOVA

Operations	SMF				SPC
	12Ori. +11Sca.	*4Ori. +11Sca.*	*4Ori. +6Sca.*	*4Ori. +1Sca.*	*64Basi sFunc.*
AutoLevel	0.9920	0.9918	0.9916	0.9912	0.9831
Blur 3x3	0.9032	0.8974	0.8885	0.8652	0.9496
Bright +10%	0.9444	0.9425	0.9431	0.9453	0.9354
Gaussian 4.0	0.9657	0.9647	0.9632	0.9584	0.9369
JPEG QF:80	0.8755	0.8720	0.8698	0.8598	0.9133
Scaling 90%	0.9801	0.9794	0.9782	0.9745	0.9515
Different Images	0.1489	0.1393	0.1427	0.1542	0.5001

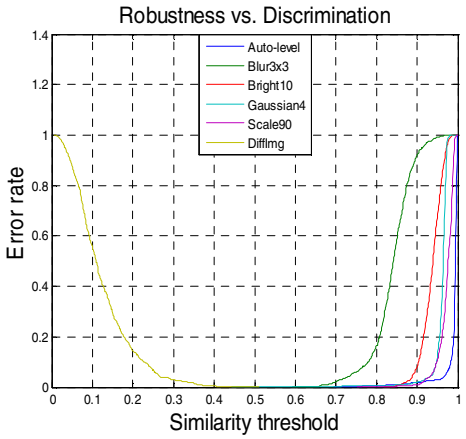

Fig. 3. Performance of SMF under 4 orientations and 1 scale.

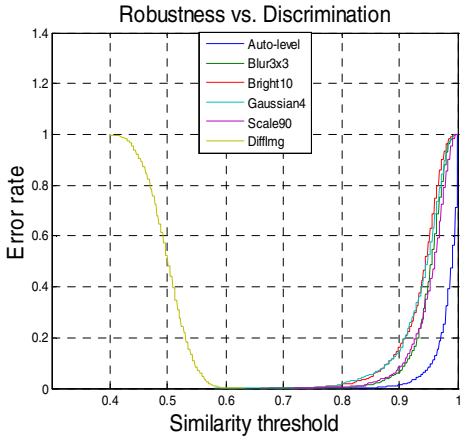

Fig. 4. Performance of SPC.

appropriate tradeoffs between robustness and discrimination so that low false alarm rates are achieved at image authentication. Here, a threshold of $\tau = 0.511$ for SMF guarantees that a false alarm rate of about 0.0003 (2 false positives out of 6020 images), with a false negative rate of 0, while $\tau = 0.633$ for SPC gives a false alarm rate of 0.00015 with a false negative rate of 0.

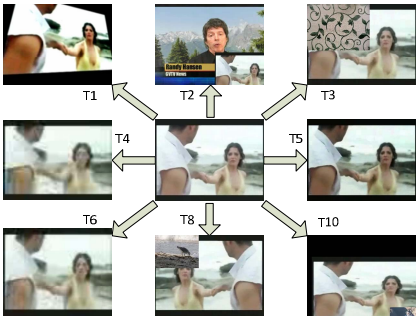

Fig. 5. Examples of video transformations.

TABLE II. AVERAGE SIMILARITIES EVALUATED OVER TRECVID

Trans.	SMF	SPC	Trans.	SMF	SPC
T3	0.4799	0.7892	T1	0.4016	0.5769
T4	0.4603	0.6870	T2	0.2741	0.5066
T5	0.5822	0.8587	T8	0.2686	0.5530
T6	0.4100	0.6743	T10	0.2673	0.5235

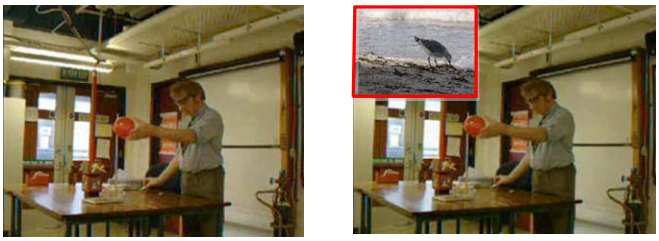

Fig. 6. An example for tamper detection and location. The tampered area is highlighted by the red bounding box.

B. *Evaluation over TRECVID*

The performance of SMF is further evaluated on the dataset derived from TRECVID [15], which contains 8 complicated transformations: simulated camcording (T1), picture in picture (T2), insertion of pattern (T3), strong re-encoding (T4), change of gamma (T5), decrease in quality (T6), post production (T8) and combined transformations (T10). Like in [2], these transformations are also roughly classified into the two categories of content-preserving and content-altering. It should be noted that the objective of image authentication is to accept moderately modified versions, while copy detection attempts to identify even severely distorted copies. In spite of this difference, SMF and SPC still show their robustness and discrimination to some extent (see Table Ⅱ). Particularly, with a Hough transform, tampered area in cases of picture in picture and pattern insertion can be successfully detected and located by extracting the SMF features from both the foreground and background and comparing the extracted SMF features against the original SMF feature (see Fig. 6).

IV. CONCLUSIONS

We propose a new invariant feature for robust and discriminating image authentication based on the biologically-plausible standard model. By exploring the characteristics of the hierarchical features, we propose to represent an image with a standard model feature (SMF), which corresponds to complex cells of cortical V1. Experimental results demonstrate that SMF possesses excellent properties of robustness and discrimination, which are comparable to those of the sparse coding feature (SPC). In the future, the feature representation and similarity measurement will be improved so that an SMF based image authentication system can be robust to more common image processing operations.

ACKNOWLEGEMENT

This work was done at National Engineering Laboratory for Video Technology, School of EE & CS, Peking University. This work was partially supported by grants from National Basic Research Program of China under contract No. 2009CB320906, Chinese National Natural Science Foundation under contract No. 60973055, National Key Technologies R&D Program of China under contract No. 2009BAH51B01 and the CADAL project.

REFERENCES

[1] A. Haouzia and R. Noumeir. Methods for image authentication: a survey. Multimedia Tools and Applications, 39(1), 1-46, Aug. 2008.

[2] L. Mou, T. Huang, Y. Tian, S. Lian, and X. Chen. Robust and discriminative image authentication based on sparse coding. Consumer Communications and Networking Conference (CCNC) , 323 –326, Jan. 2011.

[3] M. Schneider and S. Chang. A Robust Content Based Digital Signature for Image Authentication. Proc. IEEE Int'l Conf. on Image Proc.(ICIP), Vol. 3, 227-230, Sept. 1996.

[4] M. Alghoniemy, A. Tewfik. Geometric invariance in image watermarking. IEEE Trans. Image Process., 13(2), 145–153, 2004.

[5] Q. Sun and S. Chang. A Robust and Secure Media Signature Scheme for JPEG Images. Journal of VLSI Signal Processing, 41, 305-317, 2005.

[6] A. Swaminathan, Y. Mao and M. Wu. Robust and Secure Image hashing. IEEE Trans. on Information Forensics and Security, 1(2), 215-230, Jun. 2006.

[7] J. Dittmann, A. Steinmetz, and R. Steinmetz. Content-based digital signature for motion pictures authentication and content-fragile watermarking. Proceedings IEEE International Conference on Multimedia Computing and System, 2, 209–213, 1999.

[8] V. Monga and B.L. Evans. Perceptual image hashing via feature points: performance evaluation and tradeoffs. IEEE Transactions on Image Processing, 15 (11), 3453–3466, 2006.

[9] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11), 1019–1025, 1999.

[10] B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381, 607–609, 1996.

[11] T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio. A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex. AI Memo 2005-036/CBCL Memo 259, Massachusetts Inst. of Technology, Cambridge, 2005.

[12] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In CVPR, San Diego, June 2005.

[13] NOVA. http://www.amazon.co.uk/Nova-ARW-Art-Explosion-800000/dp/B0001XWNSS/ref=pd_bbs_sr_1/203-3503298-4948756?ie=UTF8&s=software&qid=1183552443&sr=8-1.

[14] Image Compression. http://www.imagecompression.info/test_images/.

[15] TRECVID. http://www-nlpir.nist.gov/projects/trecvid.